Available at
www.ComputerScienceWeb.com
POWERED BY SCIENCE DIRECT®

ELSEVIER

Theoretical Computer Science 307 (2003) 77–92

Theoretical Computer Science

www.elsevier.com/locate/tcs

Image reducing words and subgroups of free groups

D.S. Ananichev[a,1], A. Cherubini[b,*,2], M.V. Volkov[a,1,3]

[a] Department of Mathematics and Mechanics, Ural State University, 620083 Ekaterinburg, Russia
[b] Dipartimento di Matematica "Francesco Brioschi", Politecnico di Milano, 20133 Milano, Italy

Abstract

A word w over a finite alphabet Σ is said to be *n-collapsing* if for an arbitrary finite automaton $\mathscr{A} = \langle Q, \Sigma_{-.-}\rangle$, the inequality $|Q \cdot w| \leqslant |Q| - n$ holds provided that $|Q \cdot u| \leqslant |Q| - n$ for some word u (depending on $\mathscr{A}$). We give an algorithm to test whether a word is 2-collapsing. To this aim we associate to every word w a finite family of finitely generated subgroups in finitely generated free groups and prove that the property of being 2-collapsing reflects in the property that each of these subgroups has index at most 2 in the corresponding free group. We also find a similar characterization for the closely related class of so-called 2-synchronizing words.

Keywords: Synchronizing automata; Compressible automata; 2-Collapsing words; Free groups

* Corresponding author.

E-mail addresses: dmitry.ananichev@usu.ru (D.S. Ananichev), aleche@mate.polimi.it (A. Cherubini), mikhail.volkov@usu.ru (M.V. Volkov).

[1] The authors acknowledge support from Russian Education Ministry (through its Grant Center at St. Petersburg State University, Grant E00-1.0-92), from Russian Basic Research Foundation, Grant 01-01-00258, and from the INTAS (through Network Project 99-1224 "Combinatorial and Geometric Theory of Groups and Semigroups and its Applications to Computer Science").

[2] The author acknowledges support from MURST Progetto Cofinanziato 1999–2000 "Strutture Geometriche, Combinatorie e loro Applicazioni" and from MIUR Progetto Cofinanziato 2001–2002 "Linguaggi formali e automi: teoria e applicazioni".

[3] The paper was started during the author's visit to Politecnico di Milano which was supported by MURST Progetto Cofinanziato 1999–2000 "Strutture Geometriche, Combinatorie e loro Applicazioni".

0304-3975/03/$ - see front matter
doi:10.1016/S0304-3975(03)00093-8

1. Introduction

Let Q be a non-empty set. A *transformation* of the set Q is an arbitrary function f whose domain is Q and whose range (denoted by $\mathsf{Im}(f)$) is a non-empty subset of Q. The *rank* $\mathsf{rk}(f)$ of the function f is the cardinality of the set $\mathsf{Im}(f)$. For a transformation f of a finite set Q, we denote by $\mathsf{df}(f)$ its *deficiency*, that is, the difference $|Q| - \mathsf{rk}(f)$. Transformations of Q form a semigroup under the usual composition of functions; the semigroup is called *the full transformation semigroup over* Q and is denoted by $T(Q)$.

Now let Σ be a finite set called an *alphabet*. The elements of Σ are called *letters*, and strings of letters are called *words over* Σ. The set Σ^+ of all non-empty words over Σ constitutes a semigroup under the concatenation operation; the semigroup is called *the free semigroup over the alphabet* Σ. The word 'free' in this name refers to the following *universal property* of Σ^+: for every semigroup S and every map $\varphi : \Sigma \to S$, there exists a unique homomorphism of Σ^+ to S extending φ. We shall follow the common convention to let φ denote this extension homomorphism as well.

Both words over a finite alphabet and transformations of a finite set are classical objects of combinatorics. On the other hand, their interaction is essentially the main subject of the theory of finite automata. For the purpose of the present paper, a *finite automaton* $\mathscr{A}$ may be thought of as a triple $\langle Q, \Sigma, \varphi \rangle$ where Q is a finite set (called *the state set of* $\mathscr{A}$), Σ is a finite alphabet (called *the input alphabet of* $\mathscr{A}$), and φ is a mapping which assigns a transformation of the set Q to each letter in Σ. As above, φ extends to a homomorphism of the free semigroup Σ^+ into $T(Q)$ so one may speak about words over Σ acting on the state set Q via φ. For $q \in Q$ and $w \in \Sigma^+$, we shall denote the image of q under the transformation $w\varphi$ by $q \cdot w$—this simplifies the notation and should cause no confusion. For any $R \subseteq Q$, we set $R \cdot w = \{q \cdot w \mid q \in R\}$.

Let n be a positive integer. A finite automaton $\mathscr{A} = \langle Q, \Sigma, \varphi \rangle$ is said to be *n-compressible* if there is a word $w \in \Sigma^+$ such that $\mathsf{df}(w\varphi) \geqslant n$ or, in the above notation, $|Q \cdot w| \leqslant |Q| - n$. The word w is then called *n-compressing with respect to* $\mathscr{A}$. These notions (under varying names) have been around for some time mainly in connection with *Pin's conjecture* [9,10] which in our terminology can be formulated as follows: for each n-compressible automaton $\mathscr{A}$, there exists a word which is n-compressing with respect to $\mathscr{A}$ and has length n^2. Even though this particular conjecture has been recently disproved by Kari [5] (who came up with a surprising counter example in the case $n = 4$), the area—in which so easy-looking problems turn out to be so difficult to solve—remains rather vivid and reveals some interesting connections with algebra, language theory and combinatorics.

The notion that plays a central role in the present paper may be thought of as a 'black-box' version of the notion of an n-compressing word. Namely, we say that a word $w \in \Sigma^+$ is *n-collapsing* if w is n-compressing with respect to every n-compressible automaton whose input alphabet is Σ. In other terms, for $w \in \Sigma^+$ to be an n-collapsing word means that for each finite automaton $\mathscr{A} = \langle Q, \Sigma, \varphi \rangle$, we have $\mathsf{df}(w\varphi) \geqslant n$ whenever $\mathscr{A}$ is n-compressible. Thus, such a word is a 'universal tester' whose action on the state set of an arbitrary finite automaton with a fixed input alphabet exposes whether or not the automaton is n-compressible.

The very first problem related to n-collapsing words is of course the question of whether such words exist for every n. This question (which is by no means obvious) was solved in the positive by Sauer and Stone [11, Theorem 3.3] who arguably were the first to introduce words with this property (*the property* Δ_n in the terminology of Sauer and Stone [11]). While Sauer and Stone considered n-collapsing words in a purely combinatorial environment, tight relations between this notion and certain problems of automata theory (such as Pin's conjecture, for instance) were observed somewhat later by Margolis, Pin and the third-named author who extracted another existence proof from their approach [8, Theorem 2] and also found some bounds on the length of the shortest word over a given alphabet which is n-collapsing (*witnesses for deficiency* n in the terminology of Margolis et al. [8]), see [8, Theorems 5 and 11].

As the existence has been established, the next crucial step is to master, for each positive integer n, an algorithm that recognizes if a given word is n-collapsing. In Section 3 of the present paper we solve this problem for the first non-trivial case $n=2$. Our solution is based on a reduction of the initial problem to a question concerning finitely generated subgroups of finitely generated free groups. The fact that free groups intervene the area may appear a bit surprising, but as we intend to show, they do provide an adequate language for our problem. Moreover, we strongly believe that algorithms to recognize n-collapsing words for $n>2$ should also be of group-theoretic nature.

Section 4 is devoted to the closely related class of so-called 2-synchronizing words. Recall that a finite automaton $\mathscr{A}=\langle Q,\Sigma,\varphi\rangle$ is said to be *synchronizing* (or *directable*) if there exists a *reset* word $w\in\Sigma^+$ which brings all states of $\mathscr{A}$ to a particular one: $|Q\cdot w|=1$. There is an extensive body of research on synchronizing automata mainly motivated by one of the oldest open conjectures in the area—the Černý conjecture [2] that for any synchronizing automaton $\mathscr{A}$, there exists a reset word (clearly, depending on the structure of $\mathscr{A}$) of length $(|Q|-1)^2$. Being, as above, interested in a 'black-box' version of the situation, we say that a word $w\in\Sigma^+$ is *n-synchronizing* if it resets every synchronizing automaton with $n+1$ states and with the input alphabet Σ. Obviously, every n-collapsing word is also n-synchronizing but the converse is not true, see [1]. We find a characterization of 2-synchronizing words which is similar to our group-theoretic description of 2-collapsing words. The reader will see that our approach makes rather transparent what the two notions have in common and where the difference between them lies.

In Section 5 we discuss some promising directions for further research.

2. A classification of 2-compressible automata

As usual, Σ^* denotes *the free monoid* over Σ, that is, the semigroup Σ^+ with the empty word λ adjoined. If u,v are words over Σ and $u=v'vv''$ for some $v',v''\in\Sigma^*$, we say that v is a *factor* of u. It is convenient to have a name for the property of a word $w\in\Sigma^*$ to have all words of length n among its factors. We shall say that such w is *n-full*. The following is Theorem 5 in [8]:

Lemma 2.1. *If a word is n-synchronizing (in particular, if it is n-collapsing), then it is n-full.*

Now we look in some detail on 2-compressible automata. We say that a 2-compressible automaton $\mathscr{A}$ is *proper* if no word of length 2 is 2-compressing with respect to $\mathscr{A}$. The following observation is obvious:

Lemma 2.2. *If $\mathscr{A} = \langle Q, \Sigma, \varphi \rangle$ is a proper 2-compressible automaton, then for each letter $a \in \Sigma$, either $Q \cdot a = Q$ or $|Q \cdot a| = |Q| - 1$ and $Q \cdot a = Q \cdot a^2$.*

In other words, either the transformation $a\varphi$ is a permutation of Q (and then we call a a *permutation letter with respect to $\mathscr{A}$*) or $\mathsf{df}(a\varphi) = \mathsf{df}(a^2\varphi) = 1$ so that $a\varphi$ is a permutation of $Q \cdot a = \mathsf{Im}(a\varphi)$ (and then we say that a is a *just-non-permutation letter with respect to $\mathscr{A}$*). In the latter case there exists a unique state in $Q \backslash Q \cdot a$ which we call *the exception state of a* and denote by e_a. Further, since $a\varphi$ is a permutation of $Q \cdot a$, there exists a unique *doubling state* $d_a \in Q \cdot a$ such that $d_a \cdot a = e_a \cdot a$. Our next lemma, though very easy, plays an essential role in the present paper. It slightly generalizes [1, Lemma 6].

Lemma 2.3. *Let $\mathscr{A} = \langle Q, \Sigma, \varphi \rangle$ be a proper 2-compressible automaton, a, b (not necessarily different) just-non-permutation letters with respect to $\mathscr{A}$, $\alpha = a\varphi$, $\beta = b\varphi$, π an arbitrary permutation of the set Q. Then $\mathsf{df}(\alpha\pi\beta) \geqslant 2$ if and only if $e_a\pi \notin \{d_b, e_b\}$.*

Proof. It is clear that $\mathsf{df}(\alpha\pi\beta) \geqslant 2$ if and only if $d_b, e_b \in \mathsf{Im}(\alpha\pi) = \mathsf{Im}(\alpha)\pi$. Since $Q = \{e_a\} \cup \mathsf{Im}(\alpha)$ and π is a permutation, we have $Q = \{e_a\pi\} \cup \mathsf{Im}(\alpha)\pi$. Therefore the condition $d_b, e_b \in \mathsf{Im}(\alpha)\pi$ is equivalent to $e_a\pi \notin \{d_b, e_b\}$, as required. □

Given a proper 2-compressible automaton $\mathscr{A}$, let

$$\Pi_{\mathscr{A}} = \{a \in \Sigma \mid a \text{ is a permutation letter with respect to } \mathscr{A}\},$$

$$\Upsilon_{\mathscr{A}} = \{a \in \Sigma \mid a \text{ is a just-non-permutation letter with respect to } \mathscr{A}\}.$$

Proposition 2.4. *Let $\mathscr{A}$ be a proper 2-compressible automaton. Then either there exists a state e such that $e_a = e$ for all $a \in \Upsilon_{\mathscr{A}}$, or $\{d_b, e_b\} = \{d_c, e_c\}$ for all $b, c \in \Upsilon_{\mathscr{A}}$.*

Proof. Suppose that there exist two letters $b, c \in \Upsilon_{\mathscr{A}}$ such that $\{d_b, e_b\} \neq \{d_c, e_c\}$. Then the intersection $\{d_b, e_b\} \cap \{d_c, e_c\}$ contains at most one state. On the other hand, for every letter $a \in \Upsilon_{\mathscr{A}}$, its exception state e_a must belong to this intersection—otherwise by Lemma 2.3 (with the identity transformation in the role of π) one of the words ab and ac would be 2-compressing with respect to $\mathscr{A}$, in contradiction to the condition that the automaton $\mathscr{A}$ is proper. Thus, $\{d_b, e_b\} \cap \{d_c, e_c\} = e$, and $e_a = e$ for all $a \in \Upsilon_{\mathscr{A}}$. □

We call a proper 2-compressible automaton $\mathscr{A}$ a *mono automaton* whenever all letters in $\Upsilon_{\mathscr{A}}$ have a common exception state and a *stereo automaton* if there exist states $x, y \in Q$ such that $\{d_a, e_a\} = \{x, y\}$ for all $a \in \Upsilon_{\mathscr{A}}$ and there are two letters in $\Upsilon_{\mathscr{A}}$ with different exception states. In these terms, Proposition 2.4 means that each proper 2-compressible automaton is either mono or stereo.

As a corollary of Proposition 2.4, we have

Lemma 2.5. *If $\mathscr{A}$ is a proper* 2*-compressible automaton, then* $\Pi_{\mathscr{A}} \neq \emptyset$.

Proof. If $\mathscr{A}$ is a mono automaton and e is the common exception state for all letters in $\Upsilon_{\mathscr{A}}$, then, for each $w \in \Upsilon_{\mathscr{A}}^+$, we have $Q \cdot w = Q \setminus \{e\}$, whence no such w can be 2-compressing with respect to $\mathscr{A}$. Similarly, if $\mathscr{A}$ is a stereo automaton and $x, y \in Q$ are such that $\{d_a, e_a\} = \{x, y\}$ for all $a \in \Upsilon_{\mathscr{A}}$, then $Q \cdot w = Q \setminus \{x\}$ or $Q \cdot w = Q \setminus \{y\}$ for each $w \in \Upsilon_{\mathscr{A}}^+$, and again, none of such words can satisfy $|Q \cdot w| \leqslant |Q| - 2$. Since the automaton $\mathscr{A}$ is 2-compressible, we conclude that $\Sigma \neq \Upsilon_{\mathscr{A}}$, that is, $\Pi_{\mathscr{A}} = \Sigma \setminus \Upsilon_{\mathscr{A}} \neq \emptyset$. □

3. A characterization of 2-collapsing words

Our criterion involves several notions which we are going to introduce now. By a *role assignment* we shall mean an arbitrary partition of the alphabet Σ in two non-empty subsets Υ and Π. (These subsets will play the roles of $\Upsilon_{\mathscr{A}}$ and respectively, $\Pi_{\mathscr{A}}$, for the automata which will arise in our characterization.) We fix an arbitrary role assignment (Υ, Π); it will be a parameter in most of the notions we need but in order to simplify the notation we shall avoid referring to it explicitly. This should cause no confusion.

Given an arbitrary word $w \in \Sigma^+$, we can uniquely represent it in the following form:

$$w = u_0 p_1 u_1 \cdots u_{m-1} p_m u_m, \tag{1}$$

where $u_0, u_m \in \Upsilon^*$, $u_1, \ldots, u_{m-1} \in \Upsilon^+$, $p_1, \ldots, p_m \in \Pi^+$ and m is a non-negative integer. We say that the factor p_i of the decomposition (1) is an *inner segment* of the word w if both the 'neighbors' u_{i-1} and u_i of p_i are non-empty.

For a word $v \in \Sigma^*$, we denote its first and its last letter by $h(v)$ and $t(v)$, respectively. Now for each letter $a \in \Upsilon$, we define S_a to be the set of all inner factors p_i of the word w such that $h(u_i) = a$. We note that the sets S_a for different letters $a \in \Upsilon$ need not be disjoint as the same word from Π^+ may several times appear as a factor in (1) preceding different letters $a \in \Upsilon$. Let $S = \bigcup_{a \in \Upsilon} S_a$ be the set of all inner factors of w.

By $FG(\Pi)$ we denote the free group over the alphabet Π. As usual, we represent elements of $FG(\Pi)$ by words over the doubled alphabet $\{a, a^{-1} \mid a \in \Pi\}$. For $B, C \subseteq FG(\Pi)$, we define

$$B \cdot C = \{bc \mid b \in B, c \in C\} \quad \text{and} \quad B^{-1} = \{b^{-1} \mid b \in B\}.$$

Now we introduce a family of subgroups in $FG(\Pi)$. The subgroups in this family are parametrized by arbitrary subsets of the set S. Given such a subset $P \subseteq S$, we let $P_a = S_a \backslash P$ and then we define H_P to be the subgroup of $FG(\Pi)$ generated by the set $P \cup \bigcup_{a\in\Upsilon} P_a \cdot P_a^{-1}$.

Proposition 3.1. *The following conditions are equivalent for a word* $w \in \Sigma^+$:
(i) *w is 2-compressing with respect to an arbitrary mono automaton*;
(ii) *for each role assignment (Υ,Π) and for each subset $P \subseteq S$, either the subgroup H_P coincides with the group $FG(\Pi)$ or H_P has index 2 in $FG(\Pi)$ and $P_a \neq \emptyset$ for all letters $a \in \Upsilon$.*

Proof. (i)$\Rightarrow$(ii). Arguing by contradiction, suppose that for some role assignment (Υ,Π) and for some subset P of the set S of all inner segments of the word w, the subgroup H_P generated in the free group $FG(\Pi)$ by the set $P \cup \bigcup_{a\in\Upsilon} P_a \cdot P_a^{-1}$ fulfills one of the two assumptions:
(a) H_P has index >2 in $FG(\Pi)$;
(b) H_P has index exactly 2 and there exists a letter $b \in \Upsilon$ such that $P_b = \emptyset$.

In the case (a) it is possible that the index of H_P in $FG(\Pi)$ is infinite. However, H_P is generated by a finite subset of $FG(\Pi)$, and therefore, by a result due to Marshall Hall Jr. [3] H_P is equal to the intersection of those subgroups of $FG(\Pi)$ of finite index that contain H_P. This readily implies that H_P is contained in a subgroup H whose index in $FG(\Pi)$ is finite and >2. Of course, the same is true if H_P itself is of finite index as we can simply set $H = H_P$.

Departing from the subgroup H, we build an automaton $\mathscr{A}_H = \langle Q, \Sigma, \varphi\rangle$. Its state set Q is the set of all left cosets of $FG(\Pi)$ with respect to H. We note that Q is finite because the index of H is finite, and each coset in Q can be represented as Hp for some word $p \in \Pi^*$. In order to define the action $\varphi : \Sigma \to T(Q)$, we fix an arbitrary coset $H' \neq H$ and, for all $a \in \Pi$ and all $b \in \Upsilon$, set

$$Hp \cdot a = Hpa, \tag{2}$$

$$Hp \cdot b = Hp \quad \text{if } Hp \neq H, \tag{3}$$

$$H \cdot b = Hp_i, \quad \text{where } p_i \in P_b \text{ if } P_b \neq \emptyset, \tag{4}$$

$$H \cdot b = H' \quad \text{if } P_b = \emptyset. \tag{5}$$

The rule (4) may seem to be non-deterministic but it is easy to check that it is not the case. Indeed, for any $p_i, p_j \in P_b$, we have $p_i p_j^{-1} \in P_b \cdot P_b^{-1} \subseteq H$ whence $Hp_j = Hp_i p_j^{-1} p_j = Hp_i$.

By the definition, all letters $a \in \Pi$ act on Q as permutations while all $b \in \Upsilon$ are just-non-permutation letters whose common exception state is H. The automaton $\mathscr{A}_H$ is 2-compressible. Indeed, since the action of a group on its left cosets with respect to a subgroup is transitive, for an arbitrary letter $b \in \Upsilon$ there exists a word $p \in \Pi^*$ such that $Hp \notin \{H, H \cdot b\}$. By Lemma 2.3, we have $\mathsf{df}((bpb)\varphi) \geqslant 2$. We conclude that $\mathscr{A}_H$ is a mono automaton.

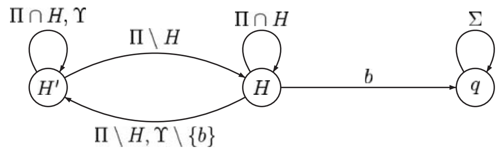

Fig. 1. The automaton $\mathscr{B}_H$.

Next we show that $\mathsf{df}(w\varphi)=1$. To this aim, we calculate step by step the images of the following transformations:

$$u_0\varphi, (u_0 p_1)\varphi, (u_0 p_1 u_1)\varphi, \ldots, \ (u_0 p_1 u_1 \cdots p_m u_m)\varphi = w\varphi. \tag{6}$$

If $u_0 \neq \lambda$ then $\mathsf{Im}(u_0\varphi)=Q\backslash\{H\}$ else $\mathsf{Im}((u_0 p_1 u_1)\varphi)=Q\backslash\{H\}$. Using this as the induction basis, assume that $\mathsf{Im}((u_0 p_1 u_1 \cdots u_{k-1})\varphi)=Q\backslash\{H\}$ for some k. Then $\mathsf{Im}((u_0 p_1 u_1 \cdots u_{k-1} p_k)\varphi)=Q\backslash\{H\cdot p_k\}$. By the rule (2) in the definition of our automaton, $H\cdot p_k = Hp_k$. If the word p_k happens to belong to the set $P \subset H$, we have $Hp_k = H = e_{h(u_k)}$. If $p_k \notin P$, then $p_k \in P_{h(u_k)}$. By rule (4), $Hp_k = H \cdot h(u_k) = d_{h(u_k)}$. Thus, in any case $Hp_k \in \{e_{h(u_k)}, d_{h(u_k)}\}$ whence $\mathsf{Im}((u_0 p_1 u_1 \cdots p_k u_k)\varphi)=Q\backslash\{H\}$, and the inductive inference takes effect. We see that the word w is not 2-compressing with respect to the mono automaton $\mathscr{A}_H$—a contradiction that shows that the case (a) is impossible.

Now consider the case (b) in which, we recall, the subgroup $H=H_P$ has index 2 and there is a letter $b\in\Upsilon$ such that $P_b=\emptyset$. Let $H'=FG(\Pi)\backslash H$; then $\{Hw \mid w\in\Pi^*\}=\{H,H'\}$. We define an automaton $\mathscr{B}_H=\langle Q,\Sigma,\varphi\rangle$ whose state set Q consists of H, H', and a new symbol q. The action $\varphi: \Sigma\to T(Q)$ is defined by the following set of rules:

$$q\cdot a = q \quad \text{for all } a\in\Sigma, \tag{7}$$

$$Hp\cdot a = Hpa \quad \text{for all } a\in\Pi, \tag{8}$$

$$H'\cdot c = H' \quad \text{for all } c\in\Upsilon, \tag{9}$$

$$H\cdot c = H' \quad \text{for all } c\in\Upsilon\backslash\{b\}, \tag{10}$$

$$H\cdot b = q. \tag{11}$$

The graphical presentation of the automaton $\mathscr{B}_H$ is shown in Fig. 1. Obviously, the letters from Υ act as just-non-permutation letters with H as a common exception state. The automaton $\mathscr{B}_H$ is 2-compressible. Indeed, since $H \subsetneq FG(\Pi)$, there exists a letter $a\in\Pi$, such that $Ha=H'$. Therefore $\mathsf{df}((bab)\varphi)\geqslant 2$ by Lemma 2.3. Hence $\mathscr{B}_H$ is a mono automaton.

As in the case (a), we show that $\mathsf{df}(w\varphi)=1$ inducting on the sequence (6) that eventually reaches $w\varphi$. The induction basis is the same as in (a). Now assume that $\mathsf{Im}((u_0 p_1 u_1 \cdots u_{k-1})\varphi)=Q\backslash\{H\}$ for some k. Then

$$\mathsf{Im}((u_0 p_1 u_1 \cdots u_{k-1} p_k)\varphi) = Q\backslash\{H\cdot p_k\}.$$

By rule (8) in the definition of our automaton, $H\cdot p_k = Hp_k \in \{H, H'\}$. If $Hp_k = H'$ then $p_k \notin P$. By the condition of the case, $S_b \setminus P = P_b = \emptyset$ whence $S_b \subseteq P$ and $p_k \notin S_b$. By the definition of the set S_b, this means that $h(u_k) \neq b$ and $Hp_k = H' = d_{h(u_k)}$. If $Hp_k = H$, then $Hp_k = e_{h(u_k)}$ because H is the exception state for all letters in Υ. Thus, we always get $H\cdot p_k \in \{e_{h(u_k)}, d_{h(u_k)}\}$ whence $\mathsf{Im}((u_0 p_1 u_1 \cdots p_k u_k)\varphi) = Q\setminus\{H\}$ completing the induction step. We again have found a mono automaton with respect to which the word w is not 2-compressing. Therefore, the case (b) is impossible too.

(ii)$\Rightarrow$(i) Take a word $w \in \Sigma^+$ satisfying (ii) and an arbitrary mono automaton $\mathscr{A} = (Q, \Sigma, \varphi)$. We aim to show that w is 2-compressing with respect to $\mathscr{A}$.

First of all, we observe that the action of the letters in Σ on the state set of $\mathscr{A}$ induces a role assignment (Υ, Π) in which $\Upsilon = \Upsilon_{\mathscr{A}}$ stands for the set of all just-non-permutation letters (obviously, it is non-empty) and $\Pi = \Pi_{\mathscr{A}}$ denotes the set of all permutation letters (it is non-empty by Lemma 2.5). Let (1) be the decomposition of the word w with respect to this particular role assignment.

Since $\mathscr{A}$ is a mono automaton, the letters in Υ share a common exception state e. Let $E = \{e\cdot p \mid p \in \Pi^*\}$ be the orbit of e under the action of the group generated by the permutations caused by the letters in Π. Clearly, this action in a natural way extends to an action of the free group $FG(\Pi)$ whence we may consider the stabilizer $ST(e)$ of the state e in $FG(\Pi)$. The index of $ST(e)$ in $FG(\Pi)$ is equal to $|E|$ because for all $\alpha, \beta \in FG(\Pi)$, the condition $e\cdot\alpha = e\cdot\beta$ is equivalent to the condition $\alpha^{-1}\beta \in ST(e)$.

Let P be the set of all those inner factors of the word w which belong to $ST(e)$. Take a letter $a \in \Upsilon$ and an inner factor $p_i \in P_a = S_a \setminus P$. By the definition of the set P, we have $e\cdot p_i \neq e$. If, besides that, $e\cdot p_i \neq d_a$ (we recall that d_a denotes the doubling state of the letter a), then by Lemma 2.3, we have $\mathsf{df}((t(u_{i-1})p_i h(u_i))\varphi) \geqslant 2$ whence $\mathsf{df}(w\varphi) \geqslant 2$. Thus, we may assume that $e\cdot p_i = d_a$ for each non-permutation letter a and for each inner factor $p_i \in P_a$, and therefore, $p_j p_i^{-1} \in ST(e)$ for all $p_i, p_j \in P_a$. We see that all generators of the subgroup H_P belong to the stabilizer $ST(e)$, that is, H_P is a subgroup of $ST(e)$. In view of the condition (ii), this implies that either $ST(e) = FG(\Pi)$ or $ST(e)$ has index 2 in $FG(\Pi)$ and $P_a \neq \emptyset$ for all letters $a \in \Upsilon$. In the case $ST(e) = FG(\Pi)$, the automaton $\mathscr{A}$ cannot be 2-compressible, in contradiction to the choice of $\mathscr{A}$. In the second case, $E = \{e, f\}$ for some state $f \neq e$. If the state f is doubling for all letters $a \in \Upsilon$, then again the automaton $\mathscr{A}$ is not 2-compressible. Therefore there exists a letter $b \in \Upsilon$ whose doubling state d_b does not belong to E.

By condition (ii), there exists an inner factor $p_i \in P_b$. Since $p_i \notin ST(e)$, we have $e\cdot p_i \in E\setminus\{e\}$, that is, $e\cdot p_i = f \notin \{e, d_b\}$. By Lemma 2.3,

$$\mathsf{df}((t(u_{i-1})p_i h(u_i))\varphi) \geqslant 2$$

whence $\mathsf{df}(w\varphi) \geqslant 2$, as required. $\square$

Now we are going to complement Proposition 3.1 by a characterization of words which are 2-compressing with respect to stereo automata. To this aim, we need another family of subgroups of the free group $FG(\Pi)$. This family arises only if the chosen role assignment (Υ, Π) satisfies $|\Upsilon| > 1$ and is parametrized by certain triples of sets (Y_1, P_{11}, P_{22}). Here Y_1 is a non-empty subset of the set Υ such that $Y_2 = \Upsilon \setminus Y_1$ is also non-empty. Further, let S_k (where $k \in \{1, 2\}$) be the set of all inner factors p_i of the word w such that $t(u_{i-1}) \in Y_k$. We note that the sets S_1 and S_2 need not be disjoint. Now we choose an arbitrary subset $P_{11} \subseteq S_1$ and an arbitrary subset $P_{22} \subseteq S_2$ and then let $P_{12} = S_1 \setminus P_{11}$ and $P_{21} = S_2 \setminus P_{22}$. The subgroup $H_{(Y_1, P_{11}, P_{22})}$ of $FG(\Pi)$ corresponding to the triple (Y_1, P_{11}, P_{22}) is generated by the set

$$P_{11} \cup P_{12} \cdot P_{21} \cup P_{12} \cdot P_{12}^{-1} \cup P_{12} \cdot P_{22} \cdot P_{12}^{-1}.$$

Proposition 3.2. *The following conditions are equivalent for a word* $w \in \Sigma^+$:

(i) *w is 2-compressing with respect to an arbitrary stereo automaton*;

(ii) *for each role assignment (Υ, Π) and for each triple (Y_1, P_{11}, P_{22}), either the subgroup $H_{(Y_1, P_{11}, P_{22})}$ coincides with the group $FG(\Pi)$ or $H_{(Y_1, P_{11}, P_{22})}$ has index 2 in $FG(\Pi)$ and $P_{12} \neq \emptyset$.*

Proof. (i)$\Rightarrow$(ii). Arguing by contradiction, suppose that for some role assignment (Υ, Π) and for some triple (Y_1, P_{11}, P_{22}), the subgroup $H_{(Y_1, P_{11}, P_{22})}$ generated in the free group $FG(\Pi)$ by the set $P_{11} \cup P_{12} \cdot P_{21} \cup P_{12} \cdot P_{12}^{-1} \cup P_{12} \cdot P_{22} \cdot P_{12}^{-1}$ is either of index > 2 or of index 2 provided that $P_{12} = \emptyset$.

As in the proof of Proposition 3.1, we substitute the subgroup $H_{(Y_1, P_{11}, P_{22})}$ by a larger subgroup H of a finite index. First suppose that $P_{12} = \emptyset$, that is, $P_{11} = S_1$. We define a finite automaton $\mathscr{C}_H = \langle Q, \Sigma, \varphi \rangle$ as follows. The state set Q of the automaton consists of all left cosets of $FG(\Pi)$ with respect to H and of one additional state q. Since the index of H is finite, the set Q is finite and each left coset can be represented as Hp for some word $p \in \Pi^*$.

The action $\varphi : \Sigma \to T(Q)$ is defined by the following set of rules:

$$\begin{array}{ll} Hp \cdot a = Hpa & \text{for all } a \in \Pi, \\ q \cdot a = q & \text{for all } a \in \Pi \cup Y_1, \\ Hp \cdot b = Hp & \text{for all } Hp \neq H \text{ and all } b \in Y_1, \\ H \cdot b = q & \text{for all } b \in Y_1, \\ Hp \cdot c = Hp & \text{for all } c \in Y_2, \\ q \cdot c = H & \text{for all } c \in Y_2. \end{array}$$

Clearly, H and q are, respectively, the exception and the doubling state for all letters $b \in Y_1$ and, at the same time, they constitute, respectively, the doubling and the exception state for all letters $c \in Y_2$. Besides that, the automaton $\mathscr{C}_H$ is 2-compressible. Indeed, since $H \neq FG(\Pi)$, there is a letter $a \in \Pi$ such that $H \cdot a = Ha \neq H$ whence by Lemma 2.3, $\mathsf{df}((bab)\varphi) \geqslant 2$ for any letter $b \in Y_1$. Thus, $\mathscr{C}_H$ is a stereo automaton. Fig. 2 represents the automaton $\mathscr{C}_H$ in the case when the index of H in $FG(\Pi)$ is equal to 2 (we denote $FG(\Pi) \setminus H$ by H').

Next we show that $\mathsf{df}(w\varphi) = 1$. Arguing by contradiction, we find the first number i such that $\mathsf{df}((u_0 p_1 u_1 \cdots p_i u_i)\varphi) \geqslant 2$. Then

$$|\mathsf{Im}((u_0 p_1 u_1 \cdots p_{i-1} u_{i-1})\varphi)| = |Q| - 1.$$

By our definition of the action of the letters in Υ, the only state in Q which does not belong to $\mathsf{Im}((u_0 p_1 u_1 \cdots p_{i-1} u_{i-1})\varphi)$ should coincide with either H or q. If this state is q, then also $\mathsf{Im}((u_0 p_1 u_1 \cdots p_{i-1} u_{i-1} p_i)\varphi) = Q \setminus \{q\}$ whence $\mathsf{df}((u_0 p_1 u_1 \cdots p_i u_i)\varphi) = 1$.

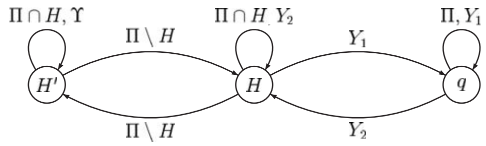

Fig. 2. The automaton $\mathscr{C}_H$ in the case $|FG(\Pi):H|=2$.

If the state in $Q\backslash \mathsf{Im}((u_0p_1u_1\cdots p_{i-1}u_{i-1})\varphi)$ is H, then $t(u_{i-1})\in Y_1$ whence $p_i\in S_1 = P_{11}\subseteq H$. Then $H\cdot p_i=Hp_i=H$ and $\mathsf{Im}((u_0p_1u_1\cdots p_{i-1}u_{i-1}p_i)\varphi)=Q\backslash\{H\}$. Thus, we again get $\mathsf{df}((u_0p_1u_1\cdots p_iu_i)\varphi)=1$.

Now suppose that $P_{12}\neq\emptyset$. This means that the index of H is at least 3. We define a finite automaton $\mathscr{D}_H=\langle Q,\Sigma,\varphi\rangle$ whose state set Q is the set $\{Hp \mid p\in\Pi^*\}$ of all left cosets of $FG(\Pi)$ with respect to H. As usual, the letters from Π act on Q in a natural way: $Hp\cdot a=Hpa$ for all $a\in\Pi$ and all cosets $Hp\in Q$. Take an inner segment $p_n\in P_{12}$ and denote the coset Hp_n by H'. Observe that for any $p_j\in P_{12}$, we have $H\cdot p_j=Hp_j=H(p_jp_n^{-1})p_n=H'$ because $p_jp_n^{-1}\in P_{12}\cdot P_{12}^{-1}\subseteq H$. Further, for any $p_k\in P_{21}$, we have $H'\cdot p_k=Hp_np_k=H$ because $p_np_k\in P_{12}\cdot P_{21}\subseteq H$. Similarly, for any $p_\ell\in P_{22}$, we have $H'\cdot p_\ell=H(p_np_\ell p_n^{-1})p_n=H'$ because $p_np_\ell p_n^{-1}\in P_{12}\cdot P_{22}\cdot P_{12}^{-1}\subseteq H$.

The action of the letters from Υ is defined by the following rules:

$$
\begin{array}{ll}
Hp\cdot b=Hp & \text{for all } Hp\neq H \text{ and for all } b\in Y_1,\\
H\cdot b=H' & \text{for all } b\in Y_1,\\
Hp\cdot c=Hp & \text{for all } Hp\neq H' \text{ and for all } c\in Y_2,\\
H'\cdot c=H & \text{for all } c\in Y_2.
\end{array}
$$

Clearly, under these rules, H and H' become, respectively, the exception and the doubling state for all letters $b\in Y_1$ and, at the same time, the doubling and the exception state for all letters $c\in Y_2$. Besides that, the automaton $\mathscr{D}_H$ is 2-compressible. Indeed, since the index of H in $FG(\Pi)$ is at least 3, there is a word $p\in\Pi^*$ such that $Hp\notin\{H,H'\}$. By Lemma 2.3, $\mathsf{df}((bpb)\varphi)\geqslant 2$ for any letter $b\in\Upsilon_1$. Hence, $\mathscr{D}_H$ is a stereo automaton.

Next we show that $\mathsf{df}(w\varphi)=1$. Arguing by contradiction, we find the first number i such that $\mathsf{df}((u_0p_1u_1\cdots p_iu_i)\varphi)\geqslant 2$. Then

$$|\mathsf{Im}((u_0p_1u_1\cdots p_{i-1}u_{i-1})\varphi)|=|Q|-1.$$

The action of the letters in Υ has been defined so that the only state $q\in Q$ which does not belong to $\mathsf{Im}((u_0p_1u_1\cdots p_{i-1}u_{i-1})\varphi)$ is either H or H'. If $q=H$, then $t(u_{i-1})\in Y_1$ whence $p_i\in S_1=P_{11}\cup P_{12}$. Therefore $H\cdot p_i=Hp_i\in\{H,H'\}$ as we have shown above. Similarly, if $q=H'$, then $t(u_{i-1})\in Y_2$ whence $p_i\in S_2=P_{21}\cup P_{22}$. Therefore $H'\cdot p_i=H'p_i\in\{H,H'\}$. In any case, the image of the transformation $(u_0p_1u_1\cdots p_{i-1}u_{i-1}p_i)\varphi$ cannot include both H and H', and since (H,H') is the only pair of states which the transformation $u_i\varphi$ glues together, we conclude that $\mathsf{df}((u_0p_1u_1\cdots p_iu_i)\varphi)=1$, a contradiction.

(ii)$\Rightarrow$(i) Take a word $w\in\Sigma^+$ satisfying (ii) and an arbitrary stereo automaton $\mathscr{A}=(Q,\Sigma,\varphi)$. We aim to show that w is 2-compressing with respect to $\mathscr{A}$.

As in the proof of Proposition 3.1, the action of the letters in Σ on the state set of $\mathscr{A}$ induces a role assignment (Υ, Π) with $\Upsilon = \Upsilon_{\mathscr{A}}$ being the set of all just-non-permutation letters and $\Pi = \Pi_{\mathscr{A}}$ being the set of all permutation letters (which is non-empty by Lemma 2.5). Let (1) be the decomposition of the word w with respect to this particular role assignment.

Since $\mathscr{A}$ is a stereo automaton, there exist two states $x_1, x_2 \in Q$ such that $\{d_a, e_a\} = \{x_1, x_2\}$ for all $a \in \Upsilon$ and the sets $Y_1 = \{b \in \Upsilon \mid e_b = x_1\}$ and $Y_2 = \{c \in \Upsilon \mid e_c = x_2\}$ are both non-empty. Let $X_k = \{x_k \cdot p \mid p \in \Pi^*\}$ $(k = 1, 2)$ be the orbit of x_k under the action of the group generated by the permutations caused by the letters in Π. (We note that if the sets X_1 and X_2 have a common state then they must coincide.) The action naturally extends to an action of the free group $FG(\Pi)$, and we consider the stabilizer $ST(x_k)$ of the state x_k in $FG(\Pi)$. The index of $ST(x_k)$ in $FG(\Pi)$ is equal to $|X_k|$ because for all $\alpha, \beta \in FG(\Pi)$, the condition $x_k \cdot \alpha = x_k \cdot \beta$ is equivalent to the condition $\alpha^{-1}\beta \in ST(x_k)$. We set $P_{kk} = ST(x_k) \cap S_k$, where, we recall, S_k denotes the set of all inner factors p_i of the word w such that $t(u_{i-1}) \in Y_k$, $k = 1, 2$.

Let $\ell = 3 - k$ and consider an arbitrary inner factor $p_i \in P_{k\ell} = S_k \backslash P_{kk}$. By the definition of P_{kk}, we have $x_k \cdot p_i \neq x_k$. If $x_k \cdot p_i \neq x_\ell$, then by Lemma 2.3, $\mathsf{df}((t(u_{i-1})p_i h(u_i))\varphi) \geqslant 2$ whence $\mathsf{df}(w\varphi) \geqslant 2$. Thus, we may assume that $x_k \cdot p_i = x_\ell$ for each inner factor $p_i \in P_{\ell\ell}$. This easily implies that

$$P_{kk} \cup P_{k\ell} \cdot P_{\ell k} \cup P_{k\ell} \cdot P_{k\ell}^{-1} \cup P_{k\ell} \cdot P_{\natural} \cdot P_{k\ell}^{-1} \subseteq ST(x_k),$$

whence the subgroup generated by the left-hand side of this inclusion is contained in $ST(x_k)$ as well. This subgroup is exactly $H_{(Y_k, P_{kk}, P_{\ell\ell})}$, thus by the condition (ii), the index of $ST(x_k)$ in $FG(\Pi)$ does not exceed 2.

If $ST(x_k) = FG(\Pi)$ for $k = 1, 2$, then we easily deduce that the automaton $\mathscr{A}$ cannot be 2-compressible. This contradicts to the choice of $\mathscr{A}$. Therefore, we may assume that the index of either $ST(x_1)$ or $ST(x_2)$ in $FG(\Pi)$ is equal to 2. By symmetry, it suffices to consider the case when $ST(x_1)$ has index 2. Then $|X_1| = 2$, that is, $X_1 = \{x_1, y\}$ for some state $y \neq x_1$. In view of the condition (ii), $ST(x_1) = H_{(Y_1, P_{11}, P_{22})}$ and $P_{12} \neq \emptyset$ whence, as we have shown, there exists an inner factor p_i of the word w such that $x_1 \cdot p_i = x_2$. Therefore $y = x_2$, an we conclude that $X_2 = X_1 = \{x_1, x_2\}$. Then the automaton $\mathscr{A}$ cannot be 2-compressible, a contradiction. $\square$

Combining Propositions 3.1 and 3.2 with Lemma 2.1, we immediately obtain the following characterization of 2-collapsing words:

Theorem 3.3. *A word $w \in \Sigma^+$ is 2-collapsing if and only if w is 2-full and for each role assignment (Υ, Π), the following conditions hold*:

(I) *for each subset $P \subseteq S$, either $H_P = FG(\Pi)$ or the subgroup H_P has index 2 in $FG(\Pi)$ and $P_a \neq \emptyset$ for all letters $a \in \Upsilon$*;

(II) *for each triple (Y_1, P_{11}, P_{22}), either $H_{(Y_1, P_{11}, P_{22})} = FG(\Pi)$ or the subgroup $H_{(Y_1, P_{11}, P_{22})}$ has index 2 in $FG(\Pi)$ and $P_{12} \neq \emptyset$.*

We note that the conditions of Theorem 3.3 can be effectively verified. Indeed, it suffices to examine finitely many subgroups, each generated by a finite set of words.

Given a finite subset U of a free group, one can effectively decide whether U generates a subgroup of finite index, moreover, if it is the case, the corresponding algorithm returns the index, see [6, Proposition I.3.22]. As for the complexity of the resulting decision procedure for the property of being 2-collapsing, it requires exponential time (as the function of the length $|w|$ of the word w under examination). Indeed, it can be calculated that for sufficiently long words w and for the pairs (Υ, Π) with $|\Pi|>1$, the number $|S|$ of inner segments is of the magnitude $O(|w|)$ whence the number of subgroups of the form H_P or $H_{(Y_1,P_{11},P_{22})}$ to be inspected is of the magnitude $O(2^{|w|})$.

Example 3.1. According to Ananichev and Volkov [1, Proposition 9], the word

$$w_{27} = abc^2b^2c \cdot bca^2c^2a \cdot cab^2a^2b \cdot (abc)^2$$

of length 27 is 2-collapsing. Now we show how this can be deduced from Theorem 3.3.

Obviously, the word w_{27} is 2-full. We have to check all possible role assignments (Υ, Π) of the alphabet $\Sigma = \{a,b,c\}$. First consider the assignment $\Upsilon = \{a\}$, $\Pi = \{b,c\}$. Decomposing the word w_{27} with respect to this role assignment, we find that the set $S = S_a$ of its inner segments consists of the following 6 words: b, c, b^2, bc, c^2, bc^2b^2cbc. Rather than going through all 64 subsets $P \subseteq S$ with the algorithm mentioned above, we come up with a short ad hoc argument. Let H stand for the subgroup H_P generated by $P \cup P_a \cdot P_a^{-1}$ where $P_a = S_a \backslash P = S \backslash P$.

If $b,c \in P$, then $H = FG(b,c)$, and we are done. Suppose that $b \in P$ and $c \in S\backslash P$. If $c^2 \in S\backslash P$, then $c = c^2 \cdot c^{-1} \in H$, and again $H = FG(b,c)$. Thus, we assume that $c^2 \in P$. Here we consider two subcases: $bc^2b^2cbc \in P$ or $bc^2b^2cbc \in S\backslash P$. In the latter subcase $bc^2b^2cb = bc^2b^2cbc \cdot c^{-1} \in H$, and since $b, c^2 \in H$, we again conclude that $c \in H$ and $H = FG(b,c)$. In the former subcase we obtain $cbc \in H$, whence also $c^{-1}bc = (c^2)^{-1} \cdot cbc \in H$. Using this and the assumption that $c^2 \in H$, it is easy to deduce that H contains the subgroup of index 2 consisting of all words $\prod_i b^{\beta_i}c^{\gamma_i} \in FG(b,c)$, $\beta_i, \gamma_i \in \mathbb{Z}$, with the sum $\sum_i \gamma_i$ being even.

Now suppose that $c \in P$ and $b \in S\backslash P$. Arguing as in the previous paragraph, we may assume that $b^2 \in P$. If $bc \in P$, then $b \in H$ and $H = FG(b,c)$. Therefore we may also assume that $bc \in S\backslash P$ whence $bcb^{-1} \in H$. This and the assumption that $b^2 \in H$ imply that H contains the subgroup of index 2 consisting of all words $\prod_i b^{\beta_i}c^{\gamma_i} \in FG(b,c)$, $\beta_i, \gamma_i \in \mathbb{Z}$, with the sum $\sum_i \beta_i$ being even.

Finally, suppose that $b,c \in S\backslash P$. Then $bc^{-1} \in H$. If $c^2 \in S\backslash P$, then $c = c^2 \cdot c^{-1} \in H$ and $b = bc^{-1} \cdot c \in H$ whence $H = FG(b,c)$. Thus, we may assume that $c^2 \in P$ and, by the same argument, $b^2 \in P$. Now it is easy to see that H contains the subgroup of index 2 consisting of all words of even length.

Thus, we have verified that either $H = FG(b,c)$ or H has index 2, and the latter situation is only possible when $P_a \neq \emptyset$. Therefore condition (I) of Theorem 3.3 holds for the role assignment $(\{a\},\{b,c\})$. The high symmetry of the word w_{27} ensures that the same arguments apply to the two other role assignments with $|\Upsilon| = 1$: the corresponding sets of inner segments differ from the above set S only by the names of the letters involved.

Now consider role assignments (Υ, Π) with $|\Upsilon| = 2$. Again, by symmetry, it suffices to analyze one of the three possible cases, for instance, $(\{a,b\},\{c\})$. Here $S = S_a = S_b = \{c, c^2\}$, and for any $P \subseteq S$, the subgroup H_P is easily seen to satisfy (I). Since $|\Upsilon| > 1$, we should also verify condition (II). If $Y_1 = \{a\}$, $Y_2 = \{b\}$, then $S_1 = S_2 = \{c, c^2\}$ and it is easy to verify that $H_{(Y_1,P_{11},P_{22})} = FG(c)$ for all choices of the subsets $P_{11}, P_{22} \subseteq \{c, c^2\}$ unless $P_{11} = P_{22} = \{c^2\}$ in which case $H_{(Y_1,P_{11},P_{22})}$ is generated by the word c^2 and has index 2 while $P_{12} \neq \emptyset$. Thus, (II) is satisfied, and the same reasoning applies when $Y_1 = \{b\}$, $Y_2 = \{a\}$.

Example 3.2. Consider the word

$$v_{27} = abcb^2c^2 \cdot bcac^2a^2 \cdot caba^2b^2 \cdot (abc)^2$$

of length 27. In spite of its similarity to the word w_{27} of Example 3.1, it is not 2-collapsing.

Indeed, consider the role assignment $\Upsilon = \{a\}$, $\Pi = \{b,c\}$. The set $S = S_a$ of inner segments with respect to this role assignment consists of the words b, c, b^2, bc, c^2, bcb^2c^2bc. Let $P = \{b, b^2, c^2, bcb^2c^2bc\}$. It can be easily verified that the subgroup H_P (which is freely generated by the words b, c^2, $cb^2c^2bc^{-1}$) has infinite index in $FG(b,c)$. Thus, condition (I) fails.

4. A characterization of 2-synchronizing words

We are going to present a group-theoretical characterization of 2-synchronizing words in the flavour of Theorem 3.3. We should mention that the property of being n-synchronizing can in principle be recognized in a direct way for any n. Indeed, there exist only finitely many automata with $n+1$ states and with a fixed input alphabet Σ, and there is an algorithm which recognizes whether or not a given finite automaton is synchronizing (see, for example, [13, Theorem 9.9]). Hence, given a word $w \in \Sigma^+$, one can straightforwardly check if w resets every synchronizing automaton over Σ with $n+1$ states. Moreover, the language of all n-synchronizing words over Σ is regular [1, Proposition 3]. It is clear, however, that this decidability in principle does not lead to any practical algorithm even for small values of n. Besides that, characterizing 2-synchronizing words in group-theoretic terms clarifies the relations between these words and the 2-collapsing ones.

Theorem 4.1. *A word $w \in \Sigma^+$ is 2-synchronizing if and only if w is 2-full and for each role assignment (Υ, Π), the following conditions hold*:

(I′) *for each subset $P \subseteq S$, the subgroup H_P is not contained in any subgroup of index 3 in $FG(\Pi)$ and if H_P is contained in a subgroup of index 2 in $FG(\Pi)$ then $P_a \neq \emptyset$ for all letters $a \in \Upsilon$;*

(II′) *for each triple (Y_1, P_{11}, P_{22}), the subgroup $H_{(Y_1,P_{11},P_{22})}$ is not contained in any subgroup of index 3 in $FG(\Pi)$ and if $H_{(Y_1,P_{11},P_{22})}$ is contained in a subgroup of index 2 in $FG(\Pi)$ then $P_{12} \neq \emptyset$.*

Theorem 4.1 readily follows from the above proof of Theorem 3.3. Indeed, if the subgroup H_P is contained in a subgroup H of index 3, we can construct the mono automaton $\mathscr{A}_H$ as in the proof of Proposition 3.1 which has 3 states, and therefore, is synchronizing while the word w does not reset it. If H_P is contained in a subgroup H of index 2 but $P_b = \emptyset$ for some letter $b \in \Upsilon$, the 3-state mono automaton $\mathscr{B}_H$ does the job. Thus, the condition (I′) is necessary. Similarly, we can use the automata $\mathscr{C}_H$ and $\mathscr{D}_H$ in order to show that also the condition (II′) is necessary. Finally, the fact that w must be 2-full follows from Lemma 2.1.

Conversely, if we start the reasoning in the proof of the implication (ii) $\Rightarrow$ (i) of Proposition 3.1 by considering an arbitrary mono automaton $\mathscr{A} = (Q, \Sigma, \varphi)$ with $|Q| = 3$, then the stabilizer $ST(e)$ of the common exception state e has index at most 3 in the group $FG(\Pi_{\mathscr{A}})$. The proof shows that either w resets $\mathscr{A}$ or $H_P \subseteq ST(e)$. In the latter case, condition (I′) guarantees that either $ST(e) = FG(\Pi_{\mathscr{A}})$ or $ST(e)$ has index 2 in $FG(\Pi_{\mathscr{A}})$ and $P_a \neq \emptyset$ for all letters $a \in \Upsilon$. Thus we can complete the proof as in Proposition 3.1. In a similar way, the proof of the implication (ii) $\Rightarrow$ (i) of Proposition 3.2 can be adapted in order to show that under (II′) the word w resets an arbitrary stereo automaton with 3 states.

Comparing Theorems 4.1 and 3.3, we see where the difference between the two involved properties lies. For instance, if one of the subgroups corresponding to a word w has index 5 then w cannot be 2-collapsing but may well happen to be 2-synchronizing (as a subgroup of index 5 cannot be contained in a subgroup of index 2 or 3). For concrete examples of 2-synchronizing but not 2-collapsing words the reader is referred to [1] (where words called 2-synchronizing in the present paper have appeared under the name ‘3-synchronizing’).

We mention that the conditions of Theorem 4.1 can be effectively verified because every finitely generated free group has only finitely many subgroups of index 2 or 3 and the subgroups can be systematically enumerated. Thus, given a finitely generated subgroup H of a finitely generated free group, one can check if H is contained in a subgroup of index 2 or 3. Of course, the resulting algorithm requires exponential time as the function of the length of the word under inspection.

5. Conclusion

We have characterized 2-collapsing words in terms of finitely generated subgroups of finitely generated free groups thus yielding an algorithm that being presented with a word recognizes whether or not the word is 2-collapsing. The algorithm however is rather complicated and a natural direction for further research consists in investigating its possible simplifications. To start with, our criterion is obtained as a combination of conditions (I) and (II) arising, respectively, in the mono and stereo case but we do not know if the two conditions are independent. More precisely, the following question is open for all alphabets Σ with more than 2 letters:

Question 5.1. *Let $w \in \Sigma^+$ be a 2-full word which is 2-compressing with respect to any mono automaton with the input alphabet Σ. Is w 2-collapsing?*

Of course, Theorem 3.3 allows to reformulate Question 5.1 as a question concerning finitely generated subgroups of free groups. Basically, it amounts to asking if the existence of a role assignment (Υ,Π) and a triple (Y_1,P_{11},P_{22}) such that the subgroup $H_{(Y_1,P_{11},P_{22})}$ violates condition (II) always implies the existence of another role assignment (Υ',Π') and a subset P such that the subgroup H_P violates condition (I).

More generally, it seems to be worth investigating the relations between the subgroups of the form H_P (respectively $H_{(Y_1,P_{11},P_{22})}$) for a fixed initial word w and varying role assignments and index subsets (respectively triples). No concrete example which we have calculated so far excludes the possibility that we can restrict ourselves to role assignments (Υ,Π) with $|\Upsilon|=1$ in the mono case and with $|\Upsilon|=2$ in the stereo case. Thus, we formulate

Question 5.2. (1) *Suppose that condition* (I) *of Theorem* 3.3 *holds for every role assignment* (Υ,Π) *with* $|\Upsilon|=1$. *Does the condition hold for an arbitrary role assignment*?

(2) *Suppose that condition* (II) *of Theorem* 3.3 *holds for every role assignment* (Υ,Π) *with* $|\Upsilon|=2$. *Does the condition hold for an arbitrary role assignment*?

Of course, a positive answer to either part of Question 5.2 would essentially simplify our algorithm as would do the positive answer to Question 5.1.

We conclude with an observation which resulted from a fruitful discussion with Stuart Margolis whose valuable comments are gratefully acknowledged. Our theorems reduce a problem that has arisen in automata theory to certain questions related to finitely generated subgroups of free groups. In turn, the most efficient algorithms developed in the theory of finitely generated subgroups in free groups use finite automata: given a finite set W of group words one can easily build a finite inverse automaton which can be shown to depend only on the subgroup H generated by W and not the set W itself [12,7,4]. The automaton which we denote by $\mathscr{A}(H)$ can be then used to provide elegant solutions for many natural questions concerning H including those that play a crucial role for the present paper. In particular, Theorems 3.3 and 4.1 can be reformulated in terms of inverse automata of the form $\mathscr{A}(H_P)$ and $\mathscr{A}(H_{(Y_1,P_{11},P_{22})})$. Rather than extracting such reformulations from the above results, it appears to be worth looking for a 'shortcut' which would directly produce from a given word w a bunch of inverse automata that controls the properties of being 2-collapsing/2-synchronizing. Finding such a shortcut might also help in a much desirable extension of our results to a similar characterization of n-collapsing and n-synchronizing words for an arbitrary n.

References

[1] D.S. Ananichev, M.V. Volkov, Collapsing words vs. synchronizing words, in: W. Kuich, G. Rozenberg, A. Salomaa (Eds.), Developments in Language Theor., Lectures Notes in Computer Science, Vol. 2295, Springer, Berlin, Heidelberg, New York, 2002, pp. 166–174.

[2] J. Černý, Poznámka k homogénnym eksperimentom s konecnými avtomatami, Mat.-Fyz. Cas. Slovensk. Akad. Vied. 14 (1964) 208–216 (in Slovak).

[3] M. Hall, A topology for free group and related groups, Ann. Math. 52 (1950) 127–139.

[4] I. Kapovich, A. Myasnikov, Stallings foldings and subgroups of free groups, J. Algebra 248 (2002) 608–668.
[5] J. Kari, A counter example to a conjecture concerning synchronizing words in finite automata, EATCS Bull. 73 (2001) 146.
[6] R.C. Lyndon, P.E. Schupp, Combinatorial Group Theory, Springer, Berlin, Heidelberg, New York, 1977.
[7] S. Margolis, J. Meakin, Free inverse monoids and graph immersions, Internat. J. Algebra Comput. 3 (1993) 79–99.
[8] S. Margolis, J.-E. Pin, M.V. Volkov, Words guaranteeing minimal image, in: M. Ito (Ed.), Proc. III Internat. Colloq. on Words, Languages and Combinatorics, World Scientific, Singapore, 2003, in press.
[9] J.-E. Pin, Le Problème de la Synchronisation. Contribution à l'Étude de la Conjecture de Černý, Thèse de 3éme cycle, Paris, 1978 (in French).
[10] J.-E. Pin, Sur les mots synchronisants dans un automate fini, Elektron. Informationverarb. Kybernet. 14 (1978) 283–289 (in French).
[11] N. Sauer, M.G. Stone, Composing functions to reduce image size, Ars. Combin. 31 (1991) 171–176.
[12] J. Stallings, Topology of finite graphs, Inv. Math. 71 (1971) 551–565.
[13] P.H. Starke, Abstrakte Automaten, Deutscher Verlag der Wissenschaften, Berlin, 1969 (in German).